



HAL
open science

A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test

Christian Francq, Svetlana Makarova, Jean-Michel Zakoïan

► To cite this version:

Christian Francq, Svetlana Makarova, Jean-Michel Zakoïan. A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test. *Econometrics*, 2007, 142 (1), pp.312. 10.1016/j.jeconom.2007.04.003 . hal-00501792

HAL Id: hal-00501792

<https://hal.science/hal-00501792>

Submitted on 12 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

A class of stochastic unit-root bilinear processes:
Mixing properties and unit-root test

Christian Francq, Svetlana Makarova, Jean-Michel
Zakoïan

PII: S0304-4076(07)00132-7
DOI: doi:10.1016/j.jeconom.2007.04.003
Reference: ECONOM 2961

To appear in: *Journal of Econometrics*

Received date: 31 January 2006
Revised date: 30 March 2007
Accepted date: 4 April 2007

Cite this article as: Christian Francq, Svetlana Makarova and Jean-Michel Zakoïan, A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.04.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jeconom

A class of stochastic unit-root bilinear processes: mixing properties and unit-root test.

CHRISTIAN FRANCO*, SVETLANA MAKAROVA †
and
JEAN-MICHEL ZAKOÏAN‡

Abstract

A class of stochastic unit-root bilinear processes, allowing for GARCH-type effects with asymmetries, is studied. Necessary and sufficient conditions for the strict and second-order stationarity of the error process are given. The strictly stationary solution is shown to be strongly mixing under mild additional assumptions. It follows that, in this model, the standard (non-stochastic) unit-root tests of Phillips-Perron and Dickey-Fuller are asymptotically valid to detect the presence of a (stochastic) unit-root. The finite sample properties of these tests are studied via Monte Carlo experiments.

JEL classification: C22, C12, C52.

Keywords: Augmented Dickey-Fuller test, Bilinear processes, GARCH, Mixing, Phillips-Perron test, Stochastic unit-roots.

*Université Lille III, GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: christian.francq@univ-lille3.fr

†National Bank of Poland and European University at St. Petersburg, Russia. E-mail: makarova@eu.spb.ru

‡Corresponding author. GREMARS and CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr

1 Introduction

It is now recognized that many economic series display nonstationarities and nonlinearities. Empirical researchers often find standard linear models, *i.e.* with iid (independent and identically distributed) innovations, inappropriate for differenced series. For this reason, recent papers dealing with unit roots have been concerned with modeling the error term of the linear dynamics as a non-iid process. Results on estimating and testing unit roots with non-iid errors can be found in Phillips (1987), Kim and Schmidt (1993), Seo (1999), Ling and Li (2003), Ling (2004), Rodrigues and Rubia (2005) and the references therein.

Charemza, Lifshits and Makarova (2005) showed that unit-roots models with bilinear errors have interesting economic interpretations, and are empirically relevant. Following this paper, we also consider a unit-root model with bilinear errors, but our specification is different. Our model allows for stationary increments, contrary to the model by Charemza et al (2005). A natural practice, followed by Charemza et al (2005), is to test for the presence of unit roots in a first step, and then to perform specifications tests on the noise dynamics in a second step. Caution is needed, however, in the blind application of standard unit root tests in the framework of non-iid errors. Rodrigues and Rubia (2005) present numerical experiments showing that non-iid errors may cause severe distortions in conventional unit-root tests. Ling (2004) provided an example of a unit-root model with non-iid errors, namely the so-called double-autoregressive model, in which the LS estimator of the AR coefficient does not converge in law to the standard Dickey-Fuller (DF) distribution. For such models, the most commonly used unit-root tests, *i.e.* the Phillips-Perron and augmented DF tests, may not have the correct asymptotic size.

An important issue for linear models with non-iid errors thus concerns the validity of those unit-root tests. Phillips (1987) and Phillips and Perron (1988) showed that, under moment and mixing conditions on the noise process, the unit-root hypothesis can be tested using the standard DF asymptotic distribution. The main goal of this paper is to establish the validity of those standard unit-root tests for the bilinear model under consideration. This requires analyzing in detail the probability structure of the model, in particular its mixing properties.¹ Apart from the unit-root testing problem, these properties have of course independent interest.

The paper is organized as follows. The general model is presented in Section 2, and the existence of stationary solutions is studied. Under a mild additional assumption on the distribution of the

¹Mixing is one way to characterize the decrease of dependence when the variables become sufficiently far apart (see *e.g.* Davidson, 1994).

iid process, the strictly stationary solution is shown to be strongly mixing in Section 3. Section 4 is devoted to examining the validity of the Phillips-Perron and augmented DF unit-root tests in our framework. Monte Carlo experiments are presented in Section 5. Concluding remarks are given in Section 6.

2 ECM with bilinear innovations

We consider the following ECM(p)-BL(q) model

$$\begin{cases} \Delta y_t = \phi y_{t-1} + \psi_1 \Delta y_{t-1} + \cdots + \psi_p \Delta y_{t-p} + u_t, \\ u_t = (1 + b_1 u_{t-1} + \cdots + b_q u_{t-q}) \epsilon_t, \quad \epsilon_t \text{ iid } (0, \sigma_\epsilon^2) \end{cases} \quad (1)$$

where $\phi, \psi_1, \dots, \psi_p, b_1, \dots, b_q$ are real coefficients and $\sigma_\epsilon^2 > 0$. The two equations have the following interpretations. The first one is a standard error correction form of an autoregressive model, capturing the linear behaviour of the possibly $I(1)$ (when $\phi = 0$) series y_t . Nonlinear effects are introduced in the second equation through a *bilinear* specification. Bilinear models have been studied by Granger and Andersen (1978), who introduced them in the time series literature, by Subba Rao and Gabr (1984) and by many others. As will be seen, under appropriate conditions on the coefficients b_i , (u_t) is a white noise, i.e. a centered non-correlated process. It is an iid white noise when all the coefficients b_i are equal to zero (i.e. when no nonlinear effects are present in the data). Bilinearity is a very general way to introduce nonlinearity in economic series. Its main advantage over other types of nonlinearities (e.g. Threshold AR, SETAR, Exponential AR) is that it is compatible with the properties of a noise. These properties are of course shared by ARCH models. The ECM(p)-BL(q) can be seen as an alternative to the extensively studied ECM(p)-ARCH(q).

Remark 2.1 The first two conditional moments of u_t are given by

$$E(u_t | u_{t-1}, \dots) = 0, \quad \text{Var}(u_t | u_{t-1}, \dots) = (1 + b_1 u_{t-1} + \cdots + b_q u_{t-q})^2 \sigma_\epsilon^2.$$

This form of conditional variance is a particular case of the quadratic ARCH, introduced by Sentana (1995), and also a particular case of Linear ARCH (LARCH), introduced by Robinson (1991) and recently studied by Giraitis, Robinson and Surgailis (2000), Giraitis and Surgailis (2002). It is seen that the conditional variance is asymmetric: a negative shock u_{t-i} may increase the conditional variance by a larger amount than a positive shock of the same magnitude. This so-called leverage-effect property is often described as one of the main stylized facts of financial time series,

and constitutes the main motivation of numerous extensions of the standard GARCH models, in particular the EGARCH (see Nelson, 1991), the GJR (Glosten, Jagannathan and Runkle, 1994), the TARARCH (Zakoian, 1994) and the APARCH models (Ding, Granger and Engle, 1993). Many nonlinear models, for instance the Markov-switching models introduced by Hamilton (1989), or nonparametric methods (see Pagan and Schwert, 1990) can also be employed to take into account asymmetric conditional heteroscedasticity. An interesting feature of the ECM-BL model, which is transparent on the news impact curve displayed in Figure 1, is that the volatility is not minimal at zero. In other words, an increase of small positive returns may lower volatility. One can imagine that the volatility is minimal when the returns correspond to the free-risk return ($-1/b_1$ on the figure). This interpretation, as well as the leverage effect, of course require $b_1 < 0$. Finally, the volatility is not bounded away from 0, contrary to most GARCH models.

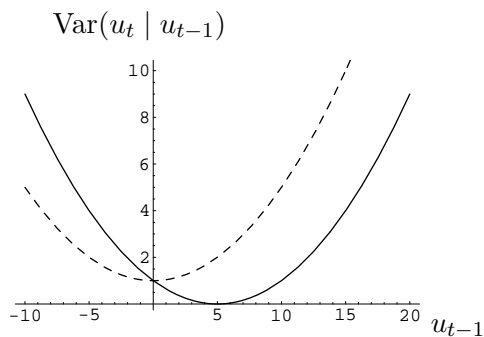


Figure 1: News impact curve of u_t in Model (1) with $q = 1$, $b_1 = -0.2$ and $\sigma_\epsilon = 1$ (full line) compared with the news impact curve of the ARCH(1) model $u_t = \sqrt{1 + b_1^2 u_{t-1}^2} \epsilon_t$ (dotted line).

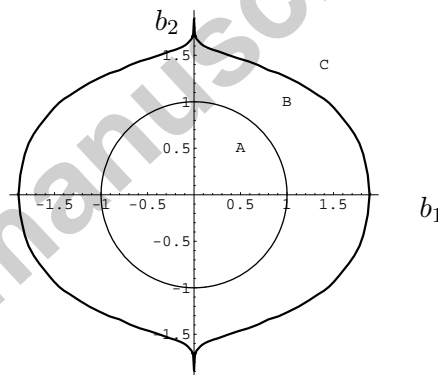


Figure 2: Strict and second-order stationarity regions of the bilinear model

$$u_t = (1 + b_1 u_{t-1} + b_2 u_{t-2}) \epsilon_t, \quad \epsilon_t \text{ iid } \mathcal{N}(0, 1)$$

A: second order stationarity, $A \cup B$: strict stationarity, and C: non stationary.

Remark 2.2 This model allows for stochastic unit-root interpretations (see Granger and Swanson (1997), Leybourne, McCabe and Tremayne (1996) for details on stochastic unit-root models). Taking $p = 0$ and $q = 1$, we have $y_t = \rho_t y_{t-1} + v_t$, where $\rho_t = 1 + \phi + b_1 \epsilon_t$ has mean 1 (justifying the name stochastic unit-root) when $\phi = 0$, and $v_t = \{1 - b(1 + \phi)y_{t-2}\} \epsilon_t$ is an error term which is uncorrelated with the y_{t-i} for $i > 0$.

Remark 2.3 For the dynamics of u_t , Charemza, Lifshits and Makarova (2005) used a slightly different bilinear specification given by

$$u_t = \epsilon_t + b \epsilon_{t-1} y_{t-1}, \quad \epsilon_t \text{ iid } (0, \sigma_\epsilon^2), \quad y_0 = \epsilon_0 = 0.$$

This model has received an economic interpretation as being derived from a model of speculative behavior. In their paper Charemza et al (2005) were mostly concerned by testing the assumption that $b = 0$, giving rise to the so-called “ b -test”. When $b \neq 0$ the error term is not stochastically stable (in particular, as demonstrated by the authors, the variance of u_t tends to infinity). Therefore this model is not suitable for $I(1)$ series.

2.1 Strict stationarity

We first give a condition for the existence of a strictly stationary white noise solution (u_t). Because, strictly speaking, (u_t) does not belong to the standard class of bilinear models², existing results, e.g. those established by Liu and Brockwell (1988), cannot be directly applied. Let $\underline{u}_t = (u_t, \dots, u_{t-q+1})' \in \mathbb{R}^q$ and $\underline{c}_t = (\epsilon_t, 0, \dots, 0)' \in \mathbb{R}^q$. Then, the second equation in (1) is equivalently written as

$$\underline{u}_t = \underline{c}_t + A_t \underline{u}_{t-1} := \begin{pmatrix} \epsilon_t \\ 0_{q-1} \end{pmatrix} + \begin{pmatrix} \mathbf{b}_{1:q-1} \epsilon_t & b_q \epsilon_t \\ I_{q-1} & 0_{q-1} \end{pmatrix} \underline{u}_{t-1}, \quad (2)$$

where $\mathbf{b}_{1:q-1} = (b_1, \dots, b_{q-1})$ and I_k is the $k \times k$ identity matrix. Notice that (\underline{c}_t, A_t) is an iid sequence of matrices. Let $\|A\| = \sum |a_{ij}|$ for any matrix $A = (a_{ij})$ and let $\log^+ x = \max(\log x, 0)$ for any positive number x . Since $E(\log^+ |\epsilon_t|) \leq E|\epsilon_t| < \infty$ we have $E(\log^+ \|A_t\|) < \infty$, and thus we can define the top-Lyapunov exponent $\gamma(\mathbf{A})$ of the sequence $\mathbf{A} = (A_t)$:

$$\gamma(\mathbf{A}) := \inf_{t \in \mathbb{N}^*} \frac{1}{t} E(\log \|A_t A_{t-1} \dots A_1\|) = \lim_{t \rightarrow \infty} a.s. \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\|. \quad (3)$$

If $\gamma(\mathbf{A}) < 0$, the unique strictly stationary solution to (2), in view of Bougerol and Picard (1992, Theorem 1.1), is

$$\underline{u}_t = \underline{c}_t + \sum_{k=1}^{\infty} A_t A_{t-1} \dots A_{t-k+1} \underline{c}_{t-k}. \quad (4)$$

It is straightforward that the strict stationarity of (\underline{u}_t) is equivalent to the strict stationarity of (u_t) . It is also seen that the strictly stationary solution is nonanticipative (*i.e.* with u_t function of the ϵ_{t-i} , $i \geq 0$) and ergodic, as a function of the iid process (ϵ_t) . By Lemma 2 given in the appendix, and Theorem 2.5 in Bougerol and Picard (1992), the sufficient condition $\gamma(\mathbf{A}) < 0$ is also necessary for the existence of a nonanticipative strictly stationary solution.

When $q = 1$ we have $\gamma(\mathbf{A}) = E \log |b_1 \epsilon_t| = \log |b_1| + E \log |\epsilon_t|$, and the strict stationarity condition $\gamma(\mathbf{A}) < 0$ takes the simpler form:

$$|b_1| < e^{-E \log |\epsilon_t|}. \quad (5)$$

²Standard bilinear models only allow terms of the form $\epsilon_{t-i} u_{t-j}$ with $i, j > 0$.

When ϵ_t is Gaussian, the necessary and sufficient condition is $|b_1|\sigma_\epsilon < 1.88736$.

When $q > 1$, the strict stationarity region can not be given explicitly. In Figure 2, the strict stationarity region has been evaluated using (3) and simulations of the sequence (A_t) in the case $q = 2$ and $\epsilon_t \sim \mathcal{N}(0, 1)$ (see Cline (2006) for methods to delineate stationarity regions in a more general framework). The strict stationarity curve passes at the points $(b_1, b_2) = (\pm e^{-E \log |\epsilon_t|}, 0)$, as can be seen from (5), and at the points $(b_1, b_2) = (0, \pm e^{-E \log |\epsilon_t|})$, as can be shown by algebraic computations. It is interesting to note that the stationarity region is not symmetric with respect to the diagonal $b_1 = b_2$.

2.2 Second-order stationarity

Results concerning the existence of second-order stationary solutions of bilinear models are well-known, and they can be straightforwardly extended to our model. Let (u_t) be a solution to the 2nd equation in (1). Then it is easily seen that $E(u_t) = 0$ and $E(u_t u_{t-h}) = E(\epsilon_t)E(1 + b_1 u_{t-1} + \dots + b_q u_{t-q})u_{t-h} = 0$ for any $h > 0$. Moreover, we have

$$\left(1 - \sum_{i=1}^q b_i^2 \sigma_\epsilon^2\right) E u_t^2 = \sigma_\epsilon^2 > 0.$$

It follows that

$$\sum_{i=1}^q b_i^2 \sigma_\epsilon^2 < 1 \tag{6}$$

is a necessary condition for second-order stationarity. It is shown in the appendix that the condition is also sufficient. Note that Giraitis, Robinson and Surgailis (2000) show that when $q = \infty$, condition (6) is necessary and sufficient for the existence of a strictly stationary and covariance stationary solution. However, as in GARCH models, strict stationarity is less restrictive than second-order stationarity when q is finite, since our condition for strict stationarity is both necessary and sufficient. Whether necessary and sufficient conditions for strict stationarity can be obtained for infinite-order models is an open issue, to our knowledge. The results of this section are summarized in the next theorem.

Theorem 2.1 *The second equation of (1) admits a strictly stationary solution (u_t) if and only if $\gamma(\mathbf{A}) < 0$, where $\mathbf{A} = (A_t)$ is defined in (2). Under this condition, the strictly stationary solution is unique, nonanticipative and ergodic. This solution admits a second order moment if and only if $\sum_{i=1}^q b_i^2 \sigma_\epsilon^2 < 1$. In this case, the solution is a conditionally heteroskedastic white noise.*

As illustrated in Figure 2, the second order stationarity region is generally much more restrictive than the strict stationarity region.

3 Mixing properties

The result of this section is stated in Theorem 3.1 below. It concerns mixing properties of the process (u_t) , which will be crucial for applying unit-root tests to Model (1). The proof relies on Markov chain techniques, and consists in showing geometric ergodicity. General conditions for ergodicity and mixing of Markov chains are provided in the book by Meyn and Tweedie (1993). References dealing with mixing properties of various classes of processes can be found in Francq and Zakoïan (2006).

Theorem 3.1 *Let f be the density of ϵ_t and assume that $f > 0$. If $\gamma(\mathbf{A}) < 0$, where $\mathbf{A} = (A_t)$ is defined in (2), then the strictly stationary solution (u_t) is strongly mixing with geometric rate.*

It is worth noting that the mixing property of Theorem 3.1 is shown without moment assumption on u_t . The proof of is given in the appendix, and relies on the following lemma.

Lemma 1 *Assume that*

- (i) (X_t) is a μ -irreducible Feller chain, for some measure μ on (E, \mathcal{E}) whose support has non-empty interior,
- (ii) (X_t) is an aperiodic chain,
- (iii) there exists a compact set $C \subset E$, an integer $m \geq 1$, and a nonnegative continuous function (test function) $g : E \rightarrow [0, +\infty)$ such that

$$\begin{aligned} E[g(X_{t+m})|X_t = x] &\leq (1 - \beta)g(x) - \beta, & x \in C^c, \\ E[g(X_{t+m})|X_t = x] &\leq b, & x \in C, \end{aligned}$$

for some strictly positive constants β and b . Then (X_t) is geometrically ergodic.

Lemma 1 is a criterion for geometric ergodicity, which has the particularity of being based on m -step transitions, instead of 1-step transitions as is usually the case. The proof is obtained from a straightforward adaptation of Meyn and Tweedie (1993, Theorem 19.1.3).

4 Unit-root testing

The Phillips-Perron and augmented DF tests are arguably the most popular unit-root tests. Both of them have been derived under precise assumptions, the validity of which is questionable for the model of this paper.

Indeed, a number of researchers (*e.g.* Gonzalo and Lee (1998), Yoon (2004) or Rodrigues and Rubia (2005)) have found, by means of numerical experiments or with theoretical arguments, that the standard unit-root tests may be in failure, or may suffer from severe size distortion or inconsistency, when the errors are not iid. The result obtained by Ling (2004) is particularly interesting: he shows that for the double autoregressive model

$$\begin{cases} y_t = y_{t-1} + \phi y_{t-1} + \epsilon_t \\ \epsilon_t = \sqrt{\omega + \alpha y_{t-1}^2} \eta_t, \end{cases} \quad (7)$$

the LSE has not the usual DF asymptotic distribution. Thus the standard unit-root tests are not valid (even asymptotically) to test $\phi = 0$ in this model, which has however an ergodic and stationary solution (under appropriate assumptions).

We start by presenting the conditions given by Phillips (1987) for the validity of the standard unit-root test, before verifying them on our model.

4.1 Phillips-Perron tests

In his seminal paper, Phillips (1987) studied the random walk

$$x_t = ax_{t-1} + v_t, \quad a = 1, \quad t = 1, 2, \dots,$$

where the initial value x_0 may be any random variable whose distribution is fixed. He showed that the standard least squares estimator $\hat{a}_n := \sum_{t=2}^n x_t x_{t-1} / \sum_{t=2}^n x_{t-1}^2$ consistently estimates $a = 1$, under very general assumptions on the error terms v_t . More precisely, denoting by $\alpha_v(k)$ the strong mixing coefficients of the process (v_t) , Phillips found that under the assumptions

- i) $E v_t = 0$ for all t ,
- ii) $\sum_{k=1}^{\infty} \{\alpha_v(k)\}^{\frac{\nu}{2+\nu}} < \infty$, for some $\nu > 0$,
- iii) $\sup_t E |v_t|^{2+\nu} < \infty$,
- iv) $\vartheta_v^2 := \lim_{n \rightarrow \infty} \text{Var} \{n^{-1/2} \sum_{t=1}^n v_t\}$ exists and $\vartheta_v^2 > 0$,

the standardized least squares estimator satisfies

$$Z_\phi := n(\hat{a}_n - 1) - \frac{n^2 \hat{\sigma}_{\hat{a}_n}^2}{2 \hat{s}_v^2} (\hat{\vartheta}_v^2 - \hat{s}_v^2) \Rightarrow \frac{(1/2) \{W^2(1) - 1\}}{\int_0^1 W^2(t) dt}, \quad (8)$$

where $\{W(t), t \in [0, 1]\}$ denotes a standard Brownian motion, $\hat{\vartheta}_v^2$ is a weakly consistent estimator of ϑ_v^2 defined in iv) above, $\hat{\sigma}_{\hat{a}_n}^2 = \hat{s}_v^2 / \sum_{t=2}^n x_{t-1}^2$, and

$$\hat{s}_v^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \hat{a}_n x_{t-1})^2 \quad (9)$$

is a weakly consistent estimator of $s_v^2 := E v_t^2$. Note that \hat{a}_n , $\hat{\sigma}_{\hat{a}_n}^2$ and \hat{s}_v^2 are available in any standard regression software. For the estimation of ϑ_v^2 , a HAC-type estimator can be used, as proposed by Phillips (1987). Phillips also found the asymptotic distribution of the associated regression t statistics:

$$Z_t := \frac{\hat{s}_v}{n \hat{\vartheta}_v \hat{\sigma}_{\hat{a}_n}} Z_\phi \Rightarrow \frac{(1/2) \{W^2(1) - 1\}}{\left\{ \int_0^1 W^2(t) dt \right\}^{1/2}}. \quad (10)$$

Note that the moment condition (iii) is not satisfied for the double autoregressive model (7). As Ling (2004) showed, the convergences (8) and (10) do not hold for this model.

4.2 Validity of the Phillips-Perron test for the bilinear process

We are interested in testing the unit-root assumption

$$H_0 : \phi = 0$$

in Model (1). We keep the notation of the previous section, with x_t replaced by y_t (and thus $v_t = y_t - y_{t-1}$). The next theorem states that (8) and (10) hold under H_0 . A drift term and/or a deterministic time trend could be added to our model, leading to the limiting distributions obtained by Phillips and Perron (1988). The stochastic unit-root hypothesis can then be tested by the standard Phillips-Perron tests, in exactly the same way as when the unit root is not stochastic.

Theorem 4.1 *Let the assumptions of Theorem 3.1 be satisfied. Assume that the zeroes of the polynomial $\psi(z) := 1 - \sum_{i=1}^p \psi_i z^i$ are outside the unit disk, and the stationary solution of the second equation in (1) satisfies $E|u_t|^{2+\nu} < \infty$ for some $\nu > 0$. Under H_0 the weak convergences (8) and (10) hold.*

The proof is given in the appendix. The estimator \hat{s}_v^2 can be replaced by the simpler estimator $n^{-1} \sum_{t=1}^n (x_t - x_{t-1})^2$. Phillips (1987, Theorem 4.2) shows that there exists a consistent HAC estimator $\hat{\vartheta}_v^2$ under the addition moment assumption $E|u_t|^{4+\nu} < \infty$. As stated in Theorem 4.1, other estimators than the HAC may be employed. The choice of the estimators of s_v^2 and ϑ_v^2 may however be important for the behavior of the statistics Z_ϕ and Z_t in finite samples and/or under the alternative $\phi \neq 0$.

For $\alpha \in (0, 1)$, let $\mathfrak{d}f_\phi(\alpha)$ and $\mathfrak{d}f_t(\alpha)$ be the α -quantiles of the distributions defined in the right-hand sides of (8) and (10). These quantiles are given in Fuller (1976, p. 371). In particular $\mathfrak{d}f_\phi(5\%) = -8.1$ and $\mathfrak{d}f_t(5\%) = -1.95$. The alternative we consider is

$$H_1 : (1 - z)\psi(z) - \phi z \neq 0 \text{ when } |z| \leq 1.$$

Under H_1 we assume that (y_t) is the nonanticipative stationary solution of (1). The following result shows, as an immediate consequence of Theorem 4.1, that the asymptotic level of the standard Phillips-Perron test remains valid in our framework. The consistency is less trivial, and is shown in the appendix.

Corollary 4.1 *We suppose that the assumptions of Theorem 4.1 are satisfied. Under the unit-root assumption H_0 ,*

$$\lim_{n \rightarrow \infty} P \{Z_\phi \leq \mathfrak{d}f_\phi(\alpha)\} = \alpha \quad \text{and} \quad \lim_{n \rightarrow \infty} P \{Z_t \leq \mathfrak{d}f_t(\alpha)\} = \alpha$$

and under the stationarity assumption H_1 ,

$$\lim_{n \rightarrow \infty} P \{Z_\phi \leq \mathfrak{d}f_\phi(\alpha)\} = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} P \{Z_t \leq \mathfrak{d}f_t(\alpha)\} = 1.$$

The last limit is obtained with the restrictions $\limsup_{n \rightarrow \infty} \hat{\vartheta}_v^2 < \infty$ a.s and $\hat{\vartheta}_v^2 > 0$ a.s for all n . The consistency of the Z_ϕ -based test is obtained whatever the nonnegative estimator $\hat{\vartheta}_v^2$.

4.3 Augmented DF tests

The approach followed by Dickey and Fuller (1979) is based on the p th-order autoregression defined by the first equation of (1):

$$\Delta y_t = (\phi, \psi')X_t + u_t, \quad \text{where} \quad X_t = (y_{t-1}, V_t)'$$

$V_t = (\Delta y_{t-1}, \dots, \Delta y_{t-p})'$ and $\psi = (\psi_1, \dots, \psi_p)'$. The least-squares estimator of $(\phi, \psi)'$ is defined by

$$(\hat{\phi}, \hat{\psi}')' = \left(\sum_{t=1}^n X_t X_t' \right)^{-1} \sum_{t=1}^n \Delta y_t X_t, \quad \hat{\psi} = (\hat{\psi}_1, \dots, \hat{\psi}_p)'$$

The following theorem is similar to Theorem 4.1-Corollary 4.1. For the sake of conciseness we only consider the test based on $\hat{\phi}$, and we omit the studentized version.

Theorem 4.2 *Assume Model (1) satisfies the assumptions of Theorem 4.1. Under H_0*

$$DF_\phi := n \frac{\hat{\phi}}{1 - \hat{\psi}_1 - \dots - \hat{\psi}_p} \Rightarrow \frac{(1/2) \{W^2(1) - 1\}}{\int_0^1 W^2(t) dt} \quad (11)$$

and, under the additional moment assumption $Eu_t^4 < \infty$,

$$\sqrt{n}(\hat{\psi} - \psi) \Rightarrow \mathcal{N}\left\{0, \Sigma_{\psi} := (EV_t V_t')^{-1} (Eu_t^2 V_t V_t') (EV_t V_t')^{-1}\right\}. \quad (12)$$

We have $\lim_{n \rightarrow \infty} P\{\text{DF}_{\phi} \leq \mathfrak{d}f_{\phi}(\alpha)\} = \alpha$ under the unit-root assumption H_0 , and $\lim_{n \rightarrow \infty} P\{\text{DF}_{\phi} \leq \mathfrak{d}f_{\phi}(\alpha)\} = 1$ under the stationarity assumption H_1 .

As in the case of an independent noise, the asymptotic null-distribution of $\sqrt{n}(\hat{\psi} - \psi)$ is the same whether the variable y_{t-1} is included or not in the regression (of course only in the case $\phi = 0$). However, the asymptotic variance Σ_{ψ} depends on the noise distribution through the b_i coefficients and the moments of ϵ (see the example below). This is not surprising because the asymptotic variance of the LS estimator in stationary ARMA models is modified when, in the noise assumptions, independence is replaced by uncorrelatedness (see Francq and Zakoïan, 1998). Interestingly, this is not the case for the distribution of $\hat{\phi}$ which turns out to be the same as for an independent noise.

In the simple case $p = 1$ with $\phi = \psi_1 = 0$, $b_1^4 < 1/3$ and $\epsilon_t \sim \mathcal{N}(0, 1)$, straightforward computations show that

$$\Sigma_{\psi} = \frac{(1 - b_1^2)(1 + 3b_1^2 + 12b_1^4)}{1 - 3b_1^4}.$$

It is seen that this asymptotic variance can be arbitrarily bigger (for b_1 close to $1/3$) than for an iid noise.

5 Small sample properties of the standard unit-root tests

This section investigates the finite-sample properties of the tests. Partial results of Monte-Carlo experiments are presented in Tables 1-2 below. Complementary results are available from the authors. In the two tables, the relative frequencies of rejection, denoted $\hat{\alpha}$, are computed over $N = 10,000$ independent replications. Table 1 displays the behavior of the tests under the unit-root hypothesis H_0 . To estimate the long-run variance ϑ_v^2 , HAC-type estimators with different kernels have been used. QS stands for the Quadratic-Spectral kernel, TH for the Turkey-Hanning kernel and Tr for the Triangular kernel (see Newey and West (1987) and Andrews (1991) for definitions). To gauge if the difference between $\hat{\alpha}$ and the nominal level α is significant or not, the statistic $z = (\hat{\alpha} - \alpha)(\alpha(1 - \alpha)/N)^{-1/2}$ is computed. Since N is large, this statistic roughly follows a standard gaussian distribution when α is the actual size of the test. A small size distortion is observed, but for very large values of b only. The difference between $\hat{\alpha}$ and α is less important with

Table 1: Empirical size: rejection relative frequencies of $H_0 : \phi = 0$ when the DGP is the URB $\Delta y_t = \phi y_{t-1} + u_t$ with $\phi = 0$ and $u_t = \epsilon_t + b\epsilon_t u_{t-1}$. The values of the z statistic are given into brackets ($z \sim \mathcal{N}(0, 1)$ if the nominal level $\alpha = 0.05$ is correct)

Statistic	b	$n = 100$	$n = 1000$
QS- Z_ϕ	0.25	0.054 (1.652)	0.050 (-0.184)
	0.99	0.075 (11.333)	0.062 (5.644)
TH- Z_ϕ	0.25	0.052 (1.055)	0.049 (-0.275)
	0.99	0.065 (6.791)	0.053 (1.193)
Tr- Z_t	0.25	0.053 (1.514)	0.051 (0.367)
	0.99	0.072 (10.048)	0.053 (1.606)
DF $_t$	0.25	0.053 (1.193)	0.051 (0.275)
	0.99	0.086 (16.656)	0.077 (12.388)

Table 2: Empirical power: rejection relative frequencies of $H_0 : \phi = 0$ when $\alpha = 0.05$, $n = 100$ and the DGP is the AR(1)-BL(1) model $y_t = ay_{t-1} + u_t$ with $u_t = \epsilon_t + b\epsilon_t u_{t-1}$.

b	$a = 0.90$		$a = 0.99$	
	Tr- Z_t	DF $_t$	Tr- Z_t	DF $_t$
0.00	0.756	0.766	0.083	0.081
0.25	0.744	0.772	0.082	0.081
0.50	0.771	0.774	0.085	0.086
0.75	0.775	0.775	0.090	0.103
0.99	0.773	0.752	0.107	0.129

the TH- Z_ϕ and Tr- Z_t tests than for the other versions. Table 2 compares the empirical powers of two tests. The powers are very close and do not vary much with b . The output of Tables 1-2, and all the other Monte Carlo experiments we performed, can be summarized as follows. The presence of bilinear terms is sensible in finite samples, however the size distortion is tiny for moderate and large sample sizes. Another teaching from our experiments is that the Phillips-Perron test performs slightly better than the augmented DF test.

6 Conclusion

In this paper we considered a class of AR models with bilinear innovations, in the spirit of Charemza et al (2005) but suitable for $I(1)$ series. This specification can be seen as a stochastic unit-root model. From another viewpoint this model is also of the GARCH type and displays asymmetries. We established necessary and sufficient strict and second-order stationarity conditions. We showed that the strict stationary solution is geometrically ergodic. Testing for unit roots in the presence of conditional heteroscedasticity is clearly important in financial applications, in particular to know if the economic shocks are persistent or not. The ergodicity results were used to demonstrate that the standard Phillips-Perron and augmented DF tests are asymptotically valid in this framework, which is not the case for other stochastic unit-roots models recently considered in the literature. Indeed, Gonzalo and Lee (1998) and Yoon (2004) showed that the standard unit-root tests do not work for detecting the stochastic unit-root hypothesis $H_0 : \phi = 0$ in the model $y_t = (1 + \phi + \alpha_t)y_{t-1} + \epsilon_t$ with $\alpha_t = \rho\alpha_{t-1} + \eta_t$, and Ling (2004) formally showed that standard unit-root tests are not asymptotically valid for the DAR model which, like the model considered in the present paper, is an AR model with GARCH-type innovations. Monte Carlo experiments have also shown that the standard Phillips-Perron and augmented DF tests have good finite sample properties for testing the stochastic unit-root hypothesis in Model (1). From these numerical experiments and the asymptotic study, we draw the conclusion that the range of application of the conventional unit-root tests is broader than the sole detection of deterministic unit-roots.

APPENDIX

We first establish a lemma, which allows to apply Theorem 2.5 in Bougerol and Picard (1992). An affine subspace H of \mathbb{R}^q is said to be invariant under (2) if it satisfies

$$\forall x \in H, \quad A_1 x + \underline{c}_1 \in H \quad \text{a.s.} \quad (13)$$

Model (2) is said to be irreducible if \mathbb{R}^q is the unique invariant affine subspace. Note that this notion of irreducibility is different from the one used in Section 3.

Lemma 2 *Model (2) is irreducible.*

Proof. For simplicity, we only give the proof for $q = 2$. The arguments are the same for $q > 2$, but the proof requires tedious notations in the general case. Let H be an affine subspace of \mathbb{R}^2 satisfying (13). By stationarity, we have, $\forall x = (x_1, x_2)' \in H$,

$$A_2(A_1 x + \underline{c}_1) + \underline{c}_2 = \begin{pmatrix} \epsilon_2(b_1^2 x_1 \epsilon_1 + b_1 b_2 x_2 \epsilon_1 + b_2 x_1 + b_1 \epsilon_1 + 1) \\ \epsilon_1(b_1 x_1 + b_2 x_2 + 1) \end{pmatrix} \in H \quad \text{a.s.} \quad (14)$$

Taking the expectation of the vector defined in (14), we obtain $0 \in H$. Taking $x = 0$ in (13) and (14), we obtain

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in H, \quad \begin{pmatrix} \epsilon_1 \\ 0 \end{pmatrix} \in H \quad \text{a.s.}, \quad \begin{pmatrix} \epsilon_2(\epsilon_1 b_1 + 1) \\ \epsilon_1 \end{pmatrix} \in H \quad \text{a.s.} \quad (15)$$

Since $\sigma_\epsilon > 0$, ϵ_1 is not almost surely equal to 0. Thus (15) entails that the linear subspace $H = \mathbb{R}^2$. \square

Proof that (6) is sufficient for the existence of a second-order stationary solution. Let the vector norm $\|X\|_2 = \{EX'X\}^{1/2}$, where X' denotes the transpose of a vector X belonging to the space L^2 of the square-integrable random variables. We will show that \underline{u}_t defined in (4) is the mean-square limit of the sequence $(\underline{u}_{tN})_N$ defined by

$$\underline{u}_{tN} = \underline{c}_t + \sum_{k=1}^N A_t A_{t-1} \dots A_{t-k+1} \underline{c}_{t-k}.$$

It suffices to show that $(\underline{u}_{tN})_N$ is a Cauchy sequence in L^2 , *i.e.* that for $N' > N$, $\|\underline{u}_{tN'} - \underline{u}_{tN}\|_2$ tends to zero when N goes to infinity. Denote by \otimes the Kronecker product of matrices and by vec the operator stacking the columns of a matrix (see *e.g.* Harville (1997) for details about these matrix operators). We have $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ and thus

$$\text{vec}(A_1 \dots A_n B C_n \dots C_1) = (C_1' \otimes A_1) \dots (C_n' \otimes A_n) \text{vec}(B) \quad (16)$$

for any conformable matrices A_i, C_i and B . It follows that, for $N' > N$,

$$\begin{aligned} \|\underline{u}_{tN'} - \underline{u}_{tN}\|_2 &\leq \sum_{k=N+1}^{N'} \|A_t \dots A_{t-k+1} \underline{c}_{t-k}\|_2 \\ &= \sum_{k=N+1}^{N'} \{E(A_t \dots A_{t-k+1} \underline{c}_{t-k})'(A_t \dots A_{t-k+1} \underline{c}_{t-k})\}^{1/2} \\ &= \sum_{k=N+1}^{N'} [E\{\text{vec}(\underline{c}'_{t-k} A'_{t-k+1} \dots A'_t A_t \dots A_{t-k+1} \underline{c}_{t-k})\}]^{1/2} \\ &= \sum_{k=N+1}^{N'} \sqrt{E \underline{c}'_t \otimes \underline{c}'_t (E A'_t \otimes A'_t)^k \text{vec} I_{q^2}}, \end{aligned} \quad (17)$$

where the inequality follows from the Minkowski inequality, and the last equality follows from (16) and the independence of the matrices A_{t-i} and \underline{c}_{t-k} . Denote by $\rho(A)$ the spectral radius of a square matrix A . Using $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$, it can be shown that the right-hand side of (17) tends to 0 as $N \rightarrow \infty$ if $\rho\{E(A_t \otimes A_t)\} < 1$. We have

$$E(A_t \otimes A_t) = \sigma_\epsilon^2 B \otimes B + J \otimes J,$$

where

$$B = \begin{pmatrix} \mathbf{b}_{1:q-1} & b_q \\ 0_{q-1 \times q-1} & 0_{q-1} \end{pmatrix}, \quad J = \begin{pmatrix} 0'_{q-1} & 0 \\ I_{q-1} & 0_{q-1} \end{pmatrix}.$$

By induction, it can be shown that

$$\det(\sigma_\epsilon^2 B \otimes B + J \otimes J - \lambda I_{q^2}) = (-\lambda)^{q^2} \mathcal{B}\left(\frac{1}{\lambda}\right),$$

where $\mathcal{B}(z) = 1 - \sum_{i=1}^q b_i^2 \sigma_\epsilon^2 z^i$. It is well-known that the roots of the polynomial $\mathcal{B}(z)$ are outside the unit disk if and only if (6) holds (see *e.g.* Francq and Zakoian (2004), Proposition 1). Thus (6) entails that the spectral radius of $E A_t \otimes A_t$ is strictly less than 1, which allows to conclude that $(\underline{u}_{tN})_N$ is a Cauchy sequence in L^2 . Therefore \underline{u}_t is in L^2 .

Proof of Theorem 3.1. To establish the geometric ergodicity of (\underline{u}_t) defined by (4) we verify the three conditions of Lemma 1. Let for $x = (x_1, \dots, x_q)' \in \mathbb{R}^q$, $\psi(x) = 1 + \sum_{i=1}^q b_i x_i$. We have

$$u_t = \psi(\underline{u}_{t-1}) \epsilon_t.$$

Let λ denote the Lebesgue measure on \mathbb{R} . For any bounded continuous function h ,

$$E(h(\underline{u}_t) | \underline{u}_{t-1} = x) = \int h(\psi(x) \epsilon, x_1, \dots, x_{q-1}) f(\epsilon) \lambda(d\epsilon)$$

is a continuous function of $x = (x_1, \dots, x_q)$, by continuity of ψ and h and by application of the Lebesgue theorem. It follows that (\underline{u}_t) is a Feller chain.

Now we will check that (\underline{u}_t) is λ_q -irreducible, where λ_q is the Lebesgue measure on $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$. To avoid cumbersome notations we will only establish this result when $q = 2$, the extension to higher dimensions being straightforward. For $B \in \mathcal{B}(\mathbb{R}^2)$ and $x = (x_1, x_2) \in \mathbb{R}^2$ we have

$$P^2(x, B) = P\{(u_2, u_1) \in B\}, \text{ where } u_1 = \epsilon_1 \psi(x), u_2 = \epsilon_2 \psi(u_1, x_1). \quad (18)$$

First consider x such that $\psi(x) \neq 0$. Let $T_x : (\epsilon_1, \epsilon_2) \mapsto (u_1, u_2)$. Let ϵ_1^0 be the point such that $\psi(u_1^0, x_1) = 0$ where $u_1^0 = \psi(x) \epsilon_1^0$. The mapping T_x is one-to-one from $(\mathbb{R} \setminus \{\epsilon_1^0\}) \times \mathbb{R}$ to $(\mathbb{R} \setminus \{u_1^0\}) \times \mathbb{R}$, and admits continuous derivatives. Since (ϵ_1, ϵ_2) admits a density, the change-of-variables theorem shows that (u_1, u_2) also admits a density. In view of (18), it follows that $P^2(x, B) > 0$ whenever $\lambda_2(B) > 0$.

Now consider x such that $\psi(x) = 0$. The previous argument fails because the distribution of $(u_1, u_2) = (0, u_2)$ has no density with respect to λ_2 . The problem is easily solved by considering three-steps transition probabilities, and by showing that (u_2, u_3) has a density whenever $\psi(0, x_1) \neq 0$. When $\psi(x) = \psi(0, x_1) = 0$, four-steps transition probabilities allow to conclude that (u_3, u_4) has a density. Hence for all x , if $\lambda_2(B) > 0$ then $P^t(x, B) > 0$ for some $t \in \{2, 3, 4\}$. This completes the proof of (i). To prove (ii) we will still limit ourselves to the case $q = 2$. Let C be a compact subset of \mathbb{R}^2 such that $\lambda_2(C) > 0$ and $\psi(x) \neq 0$ for any $x \in C$. We have just seen that, for any $x \in C$, $P^2(x, B) > 0$ whenever $\lambda_2(B) > 0$. Moreover, by continuity of the function $x \rightarrow P^2(x, B)$, the compactness of C entails that $\inf_{x \in C} P^2(x, B) = P^2(x^*, B) > 0$, for some $x^* \in C$. Setting

$\nu_2(B) = P^2(x^*, B)$, we define a non-trivial measure on $\mathcal{B}(\mathbb{R}^2)$. It follows that C is a ν_2 -small set. Now, consider the five-step transitions. We have

$$P^5(x, B) \geq \int_C P^3(x, dy) P^2(y, B) \geq P^2(x^*, B) \inf_{x \in C} P^3(x, C) = P^2(x^*, B) P^3(x^{**}, C),$$

for some $x^{**} \in C$. By arguments similar to those used in the proof of step (i), we show that $P^3(x, C) > 0$ for all $x \in C$, and thus we have $P^3(x^{**}, C) > 0$. Hence C is also ν_5 -small, with $\nu_5 = P^3(x^{**}, C)\nu_2$. In view of Meyn and Tweedie (1993, p. 116-118), we can conclude that (\underline{u}_t) is aperiodic.

Finally, we will verify condition (iii). We will use the following result which, under a slightly different form, is contained in the proof of Lemma 2.3 by Berkes, Horváth and Kokoszka (2003).

Lemma 3 *Let X be an almost surely positive random variable. If $EX^r < \infty$ for some $r > 0$ and if $E \log X < 0$, then there exists $s > 0$ such that $EX^s < 1$.*

Since $\gamma(\mathbf{A}) < 0$, there exists an integer $k > 0$ such that $E(\log \|A_t A_{t-1} \dots A_{t-k}\|) < 0$ (see the first definition of $\gamma(\mathbf{A})$ given in (3) and use the strict stationarity of the sequence (A_t)). On the other hand, we have

$$E(\|A_t A_{t-1} \dots A_{t-k}\|) \leq E\|A_t\| E\|A_{t-1}\| \dots E\|A_{t-k}\| \leq (E\|A_t\|)^{k+1} < \infty \quad (19)$$

using the facts that the norm is multiplicative and that the matrices A_t are iid. Lemma 3 entails the existence of some $s \in]0, 1[$ such that

$$\rho := E(\|A_t A_{t-1} \dots A_{t-k}\|^s) < 1. \quad (20)$$

By a recursive expansion of the first equality in (2) we get

$$\underline{u}_t = \underline{c}_t + A_t \underline{c}_{t-1} + \dots + A_t \dots A_{t-k+1} \underline{c}_{t-k} + A_t \dots A_{t-k} \underline{u}_{t-k-1}$$

and thus, the norm being multiplicative,

$$\|\underline{u}_t\| \leq \sum_{i=0}^k \|A_t \dots A_{t-i+1}\| \|\underline{c}_{t-i}\| + \|A_t \dots A_{t-k}\| \|\underline{u}_{t-k-1}\|,$$

the first term in the sum, for $i = 0$, being equal to $\|\underline{c}_t\|$ by convention. Because $s \in [0, 1)$, it follows from the elementary inequality $(a + b)^s \leq a^s + b^s$, for $a \geq 0$ and $b \geq 0$, that

$$\|\underline{u}_t\|^s \leq \sum_{i=0}^k \|A_t \dots A_{t-i+1}\|^s \|\underline{c}_{t-i}\|^s + \|A_t \dots A_{t-k}\|^s \|\underline{u}_{t-k-1}\|^s.$$

Taking the expectations in both sides, conditionally on $\underline{u}_{t-k-1} = \underline{x}$, yields

$$\begin{aligned} E(\|\underline{u}_t\|^s \mid \underline{u}_{t-k-1} = \underline{x}) &\leq \sum_{i=0}^k E\|A_t \dots A_{t-i+1}\|^s E\|\underline{c}_{t-i}\|^s + \rho \|\underline{x}\|^s \\ &\leq K + \rho \|\underline{x}\|^s. \end{aligned} \quad (21)$$

The first inequality uses the independence between the A_{t-j} and \underline{c}_{t-i} for $i > j$, and the independence between these matrices and \underline{u}_{t-k-1} for $k \geq i$. The latter independence is a consequence of the fact that the stationary solution is nonanticipative. The second inequality in (21) follows from arguments

similar to those used to show (19). Let $\beta > 0$ such that $1 - \beta > \rho$ and let C the subset of $[0, +\infty)^q$ defined by

$$C = \{\underline{x} \mid (1 - \beta - \rho)\|\underline{x}\|^s \leq K + \beta\}.$$

Clearly $C \neq \emptyset$ since $K + \beta > 0$. Moreover C is compact because $1 - \beta - \rho > 0$. Thus the right-hand side of (21) is bounded by a constant over C , and it is bounded by $(1 - \beta)\|\underline{x}\|^s - \beta$ over the complement of C . It follows that condition (iii) in Lemma 1 is verified, with $g(\underline{x}) = \|\underline{x}\|^s$, $m = k + 1$, and β and C chosen as indicated above. \square

Proof of Theorem 4.1. Note that the existence of $E|u_t|^{2+\nu}$ entails (6). First consider the case $p = 0$. Then $v_t = u_t$, and i)–iv) are straightforwardly satisfied with $\vartheta_v^2 = s_v^2 = \sigma_\epsilon^2 / (1 - \sum_{i=1}^q b_i^2 \sigma_\epsilon^2)$. Thus, when the DGP does not contain augmented variables, the result directly follows from Phillips (1987). In the case $p > 0$, it is not obvious to know whether $v_t = \psi^{-1}(B)u_t := \sum_{i=0}^{\infty} c_i u_{t-i}$ inherits the mixing property of (u_t) or not. Fortunately, conditions i)–iv) are not necessary for (8) and (10). Conditions i)–iv) are those given by Herrndorf (1984) to establish the functional central limit theorem (FCLT) for (v_t) . Other conditions ensuring the FCLT rely on the concept of near-epoch dependence (NED), see Davidson (1994). The process (v_t) is geometrically L_2 -NED on the process (u_t) because the sequence

$$\begin{aligned} \|v_t - E(v_t | u_{t-m}, \dots, u_{t+m})\|_2 &= \sum_{i=m+1}^{\infty} |c_i| \|u_{t-i} - E(u_{t-i} | u_{t-m}, \dots, u_{t+m})\|_2 \\ &\leq 2\|u_t\|_2 \sum_{i=m+1}^{\infty} |c_i| \end{aligned}$$

tends to zero at an exponential rate as $m \rightarrow \infty$. In view of this property, the exponential decrease of the α -mixing coefficients of (u_t) , and the fact that iv) holds with

$$\vartheta_v^2 = \frac{\sigma_\epsilon^2}{(1 - \sum_{i=1}^p b_i^2 \sigma_\epsilon^2) \psi^2(1)} > 0,$$

we can conclude from Corollary 29.7 in Davidson (1994), that

$$\left(\frac{1}{\sqrt{n} \vartheta_v} S_{[nt]} \right)_{t \in [0,1]} \Rightarrow (W(t))_{t \in [0,1]}, \quad (22)$$

where $S_k = v_1 + \dots + v_k$ and $[\cdot]$ denotes the integer part. As shown by Phillips (1987), (8) and (10) are direct consequences of the FCLT (22) and of the continuous mapping theorem, which completes the proof. \square

Proof of Corollary 4.1. Under H_1 we have

$$y_t = y_{t-1} + \phi y_{t-1} + \sum_{i=1}^p \psi_i \Delta y_{t-i} + u_t = \psi^{*-1}(B)u_t = \sum_{i \geq 0} \pi_i u_{t-i},$$

where $\psi^*(z) = (1 - z)\psi(z) - \phi z$. The process (y_t) is then stationary, ergodic and centered. Thus with probability one, we have

$$\hat{a}_n \rightarrow a^* := \frac{E y_t y_{t-1}}{E y_t^2} < 1,$$

where the inequality follows from the Cauchy-Schwarz inequality and the fact that the innovations of (y_t) are non degenerated. Let $\hat{v}_t^* = y_t - \hat{a}_n y_{t-1}$ and $v_t^* = y_t - a^* y_{t-1}$. The ergodic theorem also shows that

$$\begin{aligned}\hat{s}_v^2 &= \frac{1}{n-1} \sum_{t=1}^n v_t^{*2} = \frac{1}{n-1} \sum_{t=1}^n y_t^2 - \frac{2\hat{a}_n}{n-1} \sum_{t=1}^n y_t y_{t-1} + \frac{\hat{a}_n^2}{n-1} \sum_{t=1}^n y_{t-1}^2 \\ &\rightarrow s_{v^*}^2 = E v_t^{*2} = (1 - a^{*2}) E y_t^2.\end{aligned}$$

Therefore we have almost surely

$$\begin{aligned}\limsup_{n \rightarrow \infty} Z_\phi/n &= \limsup_{n \rightarrow \infty} \left\{ \hat{a}_n - 1 - \frac{1}{\frac{2}{n} \sum_{t=1}^n y_t^2} (\hat{\vartheta}_v^2 - \hat{s}_v^2) \right\} \\ &\leq a^* - 1 + \frac{s_{v^*}^2}{2E y_t^2} = -(1 - a^*) \left(1 - \frac{1 + a^*}{2} \right) < 0,\end{aligned}$$

which shows the consistency of the Z_ϕ -based test. The consistency of the Z_t -based test comes from

$$\limsup_{n \rightarrow \infty} Z_t/\sqrt{n} \leq \frac{\sqrt{E y_t^2}}{\limsup_{n \rightarrow \infty} \hat{\vartheta}_v} (a^* - 1) \left(1 - \frac{1 + a^*}{2} \right) < 0.$$

□

Proof of Theorem 4.2. We have

$$\Lambda \begin{pmatrix} \hat{\phi} \\ \hat{\psi} - \psi \end{pmatrix} = \left(\Lambda^{-1} \sum_{t=1}^n X_t X_t' \Lambda^{-1} \right)^{-1} \Lambda^{-1} \sum_{t=1}^n u_t X_t \quad (23)$$

where $\Lambda = \text{Diag}(n, \sqrt{n}, \dots, \sqrt{n})$. We have seen that the functional CLT (22) applies to $v_t := \Delta y_t = \psi^{-1}(B)u_t$. Therefore the analogue of the results (a) and (b) of Theorem 3.1 in Phillips (1987) holds. Using also the ergodic theorem, we deduce

$$\Lambda^{-1} \sum_{t=1}^n X_t X_t' \Lambda^{-1} \Rightarrow \begin{pmatrix} \frac{E u_t^2}{\psi^2(1)} \int_0^1 W^2(t) dt & 0'_p \\ 0_p & E V_t V_t' \end{pmatrix}.$$

Using Proposition 17.2 in Hamilton (1994) and the functional CLT applied to (u_t) ,

$$\begin{aligned}\frac{1}{n} \sum_{t=1}^n u_t y_{t-1} &= \frac{1}{n} \sum_{t=1}^n u_t (y_0 + v_1 + \dots + v_{t-1}) \\ &= \frac{1}{n\psi(1)} \sum_{t=1}^n u_t (u_1 + \dots + u_{t-1}) + o_P(1) \\ &\Rightarrow \frac{E u_t^2}{2\psi(1)} \{W^2(1) - 1\}.\end{aligned}$$

Moreover it is easy to show that $\hat{\psi}(1) = 1 - \hat{\psi}_1 - \dots - \hat{\psi}_p \rightarrow \psi(1)$ almost surely. The convergence (11) follows. The convergence (12) comes from the CLT applied to square integrable stationary martingale difference $(u_t V_t)$:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t V_t \Rightarrow \mathcal{N}(0, E u_t^2 V_t V_t').$$

□

Acknowledgement: We are grateful to an Associate Editor and two referees for their very helpful comments.

This work was supported by INTAS: research project number 03-51-3714.

Accepted manuscript

References

- Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Berkes, I., Horváth, L. and P.S. Kokoszka, 2003, GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.
- Bougerol, P. and N. Picard, 1992, Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.
- Charemza, W.W., Lifshits, M. and S. Makarova, 2005, Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results. *Journal of Economic Dynamics and Control*, 29, 63–96.
- Cline, D.B.H., 2006, Evaluating the Lyapounov exponent and existence of moments for threshold AR-ARCH models, *Journal of Time Series Analysis*, to appear.
- Davidson, J., 1994, *Stochastic Limit Theory*, Oxford University Press, Oxford.
- Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Ding, Z., Granger, C. and R.F. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.
- Franqc, C. and J.M. Zakoïan, 1998, Estimating Linear Representations of Nonlinear Processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Franqc, C., and J.M. Zakoïan, 2004, Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* 10, 605–637.
- Franqc, C., and J.M. Zakoïan, 2006, Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory* 22, 815–834.
- Fuller, W. A., 1976, *Introduction to Statistical Time Series*. Wiley, New York.
- Giraitis, L., Robinson, P.M. and D. Surgailis, 2000, A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability* 10, 1002–1024.
- Giraitis, L. and D. Surgailis, 2002, ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications* 100, 275–300.
- Glosten, L.R., Jagannathan, R. and D. Runkle, 1993 On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.
- Gonzalo, J. and T.H. Lee, 1998, Pitfalls in testing for long run relationships. *Journal of Econometrics* 86, 129–154.
- Granger, C.W.J. and A.P. Andersen, 1978, *An introduction to bilinear time series models*. Vandenhoeck and Ruprecht, Gottingen.
- Granger, C.W.J. and N.R. Swanson, 1997, An introduction to stochastic unit-root processes. *Journal of Econometrics* 80, 35–62.
- Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.
- Hamilton, J.D., 1994, *Time Series Analysis*. Princeton University Press, New Jersey.
- Harville, D.A., 1997, *Matrix Algebra from a Statistician’s Perspective*. Springer, New-York.
- Herrndorf, N., 1984, A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

- Kim, K. and P. Schmidt, 1993, Some evidence on the accuracy of Phillips-Perron tests using alternative estimates of nuisance parameters. *Journal of Econometrics* 59, 345–350.
- Lee, C.C. and P.C.B. Phillips, 1993, An ARMA-Prefiltered Estimator of the Long Run Variance with an Application to Unit Roots tests. Cowles Foundation Yale University, mimeographed.
- Leybourne, S.J., McCabe, B.P.M. and A.R. Tremayne, 1996, Can Economic Time Series Be Differenced to Stationarity? *Journal of Business and Economic Statistics* 14, 435–446.
- Ling, S., 2004, Estimation and testing of stationarity for double-autoregressive models. *Journal of the Royal Statistical Society: Series B* 66, 63–78.
- Ling, S. and W.K. Li, 2003, Asymptotic inference for unit root with GARCH (1,1) errors. *Econometric Theory* 19, 541–564.
- Liu, J. and P.J. Brockwell, 1988, On the general bilinear time series model. *Journal of Applied Probability* 25, 553–564.
- Meyn, S.P. and R.L. Tweedie, 1993, *Markov Chains and Stochastic stability*, 3rd Edition, Springer, London.
- Nelson, D. B., 1991, Conditional Heteroskedasticity in Asset Returns : a New Approach. *Econometrica* 59, 347–370.
- Newey, W.K. and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Pagan, A.R. and G.W. Schwert, 1990, Alternative Models for Conditional Stock Volatility. *Journal of Econometrics* 45, 267–290.
- Phillips, P.C.B., 1987, Time Series Regression with a Unit Root. *Econometrica* 55, 277–301.
- Phillips ,P.C.B. and P. Perron, 1988, Testing for a Unit Root in Time Series Regression *Biometrika* 75, 335–346.
- Quinn, B.G., 1982, A note on the existence of strictly stationary solutions to bilinear equations. *Journal of Time Series Analysis* 3, 249–252.
- Robinson, P.M., 1991, Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- Rodrigues, P.M.M. and A. Rubia, 2005, The performance of unit root tests under level-dependent heteroskedasticity. *Economics Letters* 89, 262–268.
- Seo, B., 1999, Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–144.
- Sentana, E., 1995, Quadratic ARCH models. *Review of Economic Studies* 62, 639–661.
- Subba Rao, T. and M.M. Gabr, 1984, *An introduction to bispectral analysis and bilinear time series models*. Lecture Notes in Statistics, Springer, New-York.
- Yoon, G., 2004, A re-evaluation of the performance of the unit root and cointegration tests under STUR: no more pitfalls. Department of Economics, Pusan National University, Pusan, S. Korea, preprint.
- Zakoian, J.M., 1994, Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18, 931–955.